DOCUMENT RESUME

ED 026 853 56 EM 006 968

By-Krumboltz, John D.

Factors Affecting the Design of Effective Teaching Machine Programs; January 1, 1961 - January 31, 1964. Final Report.

Stanford Univ., Calif. School of Education.
Spons Agency-Office of Education (DHEW), Washington, D.C. Educational Media Branch.
Report No-NDEA-7
Bureau No-BR-5-0851
Pub Date [31 Jan 64]
Grant-OEG-7-14-1380-190
Note-78p.
EDRS Price MF-$0.50 HC-$4.00
Descriptors-Evaluation Criteria, Intermode Differences, *Programed Instruction, Programed Materials, Programed Tutoring, Prompting, Redundancy, *Reinforcement, Research Criteria, *Response Mode, Retention

A series of studies investigating effects of alternative methods of writing, arranging, and responding in programed instruction are presented and the results are summarized to provide guidelines for use in program preparation. In a comparison of overt and covert response modes no differences were obtained on an immediate post-program criterion test; however, overt responses were found more effective on a delayed retention measure. When an overt response was required, it was more effective given as a sentence in context than as a single word. Considering the frequency and scheduling of reinforcers, increases in reinforcement decreased error rate on an immediate posttest and increased student interest in and perceived value of the program. Studies of cueing and examining procedures indicated that typographical cues improved performance when a program was difficult but had an interfering effect on moderate or easy programs and that increases in the number and plausibility of multiple-choice alternatives improved retention generally. In a review of the entire study series, it was noted that error rate during learning was not related to any retention measure and therefore appeared to be an inappropriate criterion of program adequacy. Criteria considered appropriate for program evaluation and research planning are discussed. (SS)

Factors Affecting the Design of Effective Teaching Machine Programs

John D. Krumboltz

Final Report of
Research Grant OEG 7-14-1380-190
Project No. 5-0851 II

Funded through

Educational Media Branch
U.S. Office of Education
under Title VII of the
National Defense Education Act

January 1, 1961 - January 31, 1964

School of Education
Stanford University

Stanford, California

Factors Affecting the Design of Effective Teaching Machine Programs

Final Report of

Research Grant OEG 7-14-1380-190

Project No. 5-0851 II

Funded through the

Educational Media Branch
U.S. Office of Education

under Title VII of the
National Defense Education Act

January 1, 1961 - January 31, 1964

John D. Krumboltz, Principal Investigator

School of Education
Stanford University

Stanford, California

Table of Contents

Abstract

The purpose of this project was to investigate some alternative methods of writing programed materials to enhance learning. The subject matter of the programed material was concerned with educational measurement. A series of experiments were conducted to test alternative ways of writing, arranging and responding to these materials. This abstract presents a brief summary of each experiment and refers to the chapter in which further details can be obtained. The abstract concludes with some guidelines for program writers based on extrapolations from these research findings. At the end of the report appendices are included to illustrate the nature of the materials used in one of the experiments.

Chapter I: To test the effect of overt vs. covert responding in programed instruction, 54 undergraduates in educational psychology were randomly assigned to four groups: a group who wrote down each response, a group who "mentally composed" each response, a group who read the program in which the blanks were already filled, and a control group who wrote their answers to a completely different program of about the same length. A 50-item test was administered following the study period, and an alternate form two weeks later. The three response mode groups did not differ significantly on the first test. However, on the delayed test the written response group scored significantly higher than the other two groups. The control group scored significantly lower on both tests. Thus, overt responding appears to increase delayed retention.

Chapter II: The effect of intermittent confirmation was tested on 121 students by omitting various patterns of confirming answers from a programed textbook on educational measurement. The schedules included four levels of fixed-ratio confirmation and two of variable-ratio

confirmation. Results based on criterion measures consisting of (a) errors made on the program and (b) performance on a posttest indicated: a negative linear relationship between the number of errors made on the program and the percentage of confirmation provided, no significant effects on the posttest from the various proportions of confirmation, and no evidence of differential effect between fixed-ratio and variable-ratio confirmation on either criterion.

Chapter III: An experiment was conducted comparing two approaches of presenting the confirming response in a programed textbook designed to teach prospective teachers how to write valid classroom achievement tests. The "isolation" approach consisted of presenting the desired response to the stimulus frame as a single word or phrase in the traditional programed manner. The "context" approach consisted of presenting the confirming response as a complete thought usually by inserting the desired response in a repetition of the relevant part of the stimulus frame. A control group received a completely different program. Thirty-two subjects were randomly assigned to the three treatment groups.

Although the "context" groups did not exceed the "isolation" group in knowledge of terminology on the criterion text, they did excel significantly on ability to apply principles of test construction. If confirmed by further research, this later finding would support the theoretical position that each contiguous pairing of stimulus and response strengthens the association between them and would suggest a modification in the manner of presenting the confirming response in programed material.

Chapter IV: Fifty-three undergraduates and 67 graduate students were randomly assigned to four groups: a standard key-word program group that responded with an important concept to each frame, a trivial-word program group that responded with a minor word, a paragraph-format group that read the identical material written as textbook prose, and a control group that studied and responded to a program covering different topics. Parallel forms of a criterion test were administered immediately after a two-day study period and again two weeks later. In general, both the standard key-word program group and paragraph-format group scored about the same, and both were higher than the trivial-word program group. The results cast doubt on the importance of requiring a response but emphasize the critical nature of any responses that are required.

Chapter V: Six degrees of reinforcement in a 177-frame program were provided by modifying both the number of questions asked and the number of confirming answers. Findings: (a) The more reinforcement, the fewer errors made on an immediate criterion test and on the program itself. (b) The more reinforcement, the more Ss perceived the program as interesting and a valuable learning experience. (c) When the same criterion test was delayed two months for Ss not receiving the immediate test, there was no statistically significant difference in criterion test scores among Ss subjected to various reinforcement conditions.

Chapter VI: One experiment investigated the effect of typographical cues on hard, medium, and easy programs. The medium form consisted of a simple straightforward exposition of the main points, with examples.

The easy form was identical except that additional content hints and cues were provided in order to make it more likely that the student would answer correctly. The hard form was made more difficult by complicating the sentence structure, by adding steps necessary to solve the problems, by including irrelevant information, and by using less familiar terminology. For those programs which were written without typographical cues, the hard, medium, and easy programs had difficulty levels of 22%, 7%, and 4%, respectively. For the cued programs, the error rates for the hard, medium, and easy programs were 16%, 6%, and 3%, respectively. Typographical cues reduced error rate on the program but improved immediate criterion-test performance with a diffcult program. They interfered with criterion-test performance on a medium difficulty-level program.

The second study attempted to determine the effects of two specific variations in the program: adding irrelevant information and complicating the sentence structure. On the average, the control programs had an error rate of about 8%; irrelevant programs had an error rate of about 18%; and the complex programs had an error rate of about 24%. This study failed to reveal significant differences in learning in spite of the fact that reliable differences in difficulty level had been produced either by adding irrelevant imformation or by adding complex wording and complex problems.

In the third study multiple-choice programs were constructed with variations along three dimensions: (1) the plausibility or implausibility of the alternatives; (2) the number of alternatives provided,

either two or four; and (3) the form of the answer, either recording the letter of the correct alternative or writing out the correct alternative in its entirety. While none of the differences in this study proved to be very large, it was interesting to note a consistent thread throughout all of them. The condition that produced the higher error rate also tended to produce better performance on the criterion tests. Thus, for example, providing more plausible alternatives on the program caused students to make more errors on the program but tended to improve their performance on the delayed constructed-response criterion test. Requiring them to discriminate among four alternatives rather than two produced a higher error rate but increased their performance on the immediate multiple-choice criterion test. Asking students to write out a complete answer rather than merely writing the letter of the correct answer caused students to perform better on the multiple-choice immediate criterion test even though it did not affect their error rate on the program.

Chapter VII: A program can be judged as adequate when all members of the target population demonstrate complete mastery of all the desired behaviors at the time desired by the program writer. To begin work toward this standard, we must first (a) state the terminal behaviors we desire, (b) state when they will be needed, and (c) have a procedure for assessing the degree to which each behavior has been mastered.

A perfectly adequate program will seldom be produced, so perfection must inevitably be compromised. But the way in which it is to be compromised is of critical importance. One may wish to reduce the percent of the target population that one hopes to reach; one may reduce the

percent of skills and concepts to be communicated; one may reduce the percent of mastery to be attained; or one may reduce the time over which learning is to be retained. Certainly, the time and energy which students and program writers can devote to the program puts realistic limits on what can be accomplished. Some of these compromises may be quite permissible while others would have adverse consequences.

Inadequate and insufficient bases for evaluation include program error rate, reputation of author or publisher, student opinion, and number of revisions or number of experimental subjects. Complete and timely mastery of all desired behaviors remains the ultimate standard for evaluating instruction.

Chapter VIII: Experimentation in programed instruction should not be devoted to comparisons between programed and conventional instruction. Instead, experiments would be more beneficial if they were centered on four fundamental types of questions: (1) How can programs be written to attain the most important objectives of education? (2) How can optimal curriculum sequences be determined through experimentation in programed instruction? (3) What improvements in teaching methods can be derived from research in programed instruction? (4) What modifications in instructional materials are needed to adapt them to the individual differences of pupils?

Chapter IX: Four sequences of rule and example frames were designed to compare inductive and deductive methods in programed instruction; each method had both a high and a low frequency of alternation between rules and examples. A total of 272 undergraduates responded to a 117-frame program on test interpretation in a mean time of 100 minutes.

The main findings were as follows:

1. The inductive group made more errors on the program, took less time to answer test questions on rules, but liked their method of instruction less than the deductive group liked theirs;

2. Neither method of teaching nor frequency of alternation produced significant differences in scores on the criterion test;

3. Interaction effects with intelligence appeared only when the criteria were amount of time taken to answer the test questions and number of errors made on the program;

4. Correlations between number of program errors and test scores were low and negative.

Guidelines for Program Writers

Although generalizations are dangerous, an attempt will be made here to extrapolate from the results of these experiments and formulate some advice to program writers. It must be remembered that all findings were based on one type of subject matter (educational measurement) in one style of programing (linear) with a limited range of subjects (high school and college students).

If the conclusions of these experiments are confirmed by other research, the following advice may be warranted:

1. A program may not be needed. Sophisticated subjects learning only moderately difficult material may be able to learn just as well and more quickly from conventional text material.

2. If you want subjects to retain what they learn, insist that they <u>write</u> their responses to the programed material.

3. Provide reinforcing answers for every frame. Although no differences between partial and continuous reinforcement occur on criterion tests, students make fewer errors on the program itself and rate the program as more interesting and valuable under continuous reinforcement.

4. Put reinforcing answers in context. Do not merely present the answer as one word of number in isolation, but include it in a sentence.

5. Require that the student supply a critical, not a trivial, answer. If the student can get the correct answer without being required to engage in the desired thought processes, he will not attain the educational goal as well.

6. In multiple-choice type programed materials include more than two alternatives and make the incorrect alternatives sound plausible.

7. Use typographical cues sparingly and only in the most difficult programs. In moderately difficult programs they may actually inhibit learning.

8. Don't assume that the error rate on the program itself indicates how well students will learn from it. Adding irrelevant information and requiring more complex reading and problem-solving increase error rate but not criterion performance. Additional cues and hints decrease error rate but not criterion performance.

9. If you want students to express liking for the program, sequence frames so that generalizations are presented prior to examples. However, amount of learning on criterion tests appears unaffected by the variations in sequence tried in the experiments.

10. Consider a program adequate when all members of the target population are able to emit all the desired behaviors for an appropriate length of time after instruction.

Chapter I

The Effect of Overt Versus Covert Responding To Programed Instruction on Immediate And Delayed Retention

by

John D. Krumboltz and Ronald G. Weisman

Originally published in the Journal of Educational Psychology, 1962, Vol. 53, No. 2, 89-92.

Chapter II

The Effect of Intermittent Confirmation In Programed Instruction

by

John D. Krumboltz and Ronald G. Weisman

Originally published in the Journal of Educational Psychology, 1962, Vol. 53, No. 6, 250-253.

Chapter III

The Effect of Receiving the Confirming Response In Context in Programmed Material

by

John D. Krumboltz and Barbara Bonawitz

Originally published in the Journal of Educational Research, 1962, Vol. 55, No. 9,, 472-475.

Chapter IV

The Nature and Importance of the Required Response in Programed Instruction

by

John D. Krumboltz

Originally published in the American Educational Research Journal, 1964, Vol. 1, No. 4, 203-209.

Chapter V

The Partial Reinforcement Paradigm And Programed Instruction

by

John D. Krumboltz and Charles A. Kiesler

Originally published in the Journal of Programed Instruction, 1965, Vol. 3, No. 2, 9-14.

Factors Affecting Difficulty Level and Criterion Performance

John D. Krumboltz and David Rawnsley

One persistent controversy in the field of programed learning has concerned the optimum difficulty level of the material to be learned. Skinner has consistently advocated error-free learning, while others, Pressey among the most articulate, have pointed out not only that it is possible to learn by making errors but that it is often more efficient too. Since the controversy concerned a question where empirical evidence would be relevant, an experiment seemed a logical step to settle the controversy. Why not do the crucial experiment that would settle once and for all the effect of difficulty level on subsequent learning and retention?

The problem seemed simple at first. All one would have to do would be to construct programs of different difficulty levels, randomly assign subjects to the different programs, and observe their criterion performance. The simplicity of this study disappears immediately. How do you write programs of different difficulty levels? The difficulty level of programed materials is a function of many factors and cannot be manipulated directly.

Difficulty level is a dependent variable, not an independent variable, and may vary directly or inversely with criterion performance depending on a number of independent variables which influence both difficulty level and criterion performance.

All experiments were conducted with meaningful learning material

concerned with educational measurement. In general, programs were written to teach teachers how to interpret educational and psychological test scores. Topics covered included means and grade equivalent scores, medians and percentile scores, the standard deviation, the normal curve, validity and correlation. The general plan of all the studies involved constructing materials corresponding to the main independent variables, randomly assigning subjects to the one, two or three dimensional factorial designs, and measuring performance once or twice thereafter.

The first experiment in this series was similar to one that James Holland reported to the American Psychological Association three years ago. Two forms of a linear constructed response program were prepared. One form, called the Key-word form, asked students to fill in some critical word or number which indicated that he had understood the main point of the frame. The second form, called the Trivial-word form, asked the student to fill in some minor preposition or other word which would be obvious from the sentence structure even to people who did not understand the main point of the frame. Both forms were identical in all other respects. The study was performed twice, once with graduate students and once with undergraduates. The results were virtually identical for both groups. On the average the error rate on the program was lower in the Trivial-word form (4.2%) than in the Key-word form (6.3%). However, students taking the Key-word form produced significantly higher learning both on criterion tests administered two days after learning and also on an alternate form criterion test administered two weeks after learning. The main conclusion is that reducing

the difficulty level of a program by asking for trivial responses on the program reduces the amount of learning also.

The second experiment was designed to answer some rather obvious criticisms of the first study. After all, no one has ever advocated that a program ask for trivial responses. What would happen if everyone were asked to make the same critical "key-word" response but were given various kinds of cues to suggest the correct response? Three forms of the program, labeled Hard, Medium and Easy, were constructed. The Medium form consisted of a simple straightforward exposition of the main points with examples. The Easy form was identical except that additional content hints and cues were provided in order to make it more likely that the student would answer correctly. The Hard form was similar in that it still called for the same answer as the other two forms and provided the same information, but obtaining the correct answer was made more difficult by complicating the sentence structure, by adding steps necessary to solve the problems, by including irrelevant information, and by using less familiar terminology. So now we had programs written at three levels of difficulty. We added a second dimension to this study by adding typographical cues to half the programs at each difficulty level. The typographical cues consisted of such things as giving the first and/or last letters of the correct answer, and providing underlining to indicate the number of letters in the correct answer. The error rate on the program itself indicated clearly that we had been successful in producing programs of different difficulty levels. For those programs which were written without typographical cues, the Hard, Medium, and Easy programs had difficulty

levels of 22 percent, 7 percent, and 4 percent, respectively. For the cued programs, the error rates for the Hard, Medium and Easy programs were 16 percent, 6 percent, and 3 percent, respectively. Thus, in general, the addition of typographical cues significantly reduced the number of errors made on the program. The use of various content hints and complications was also successful in producing programs with significantly differing levels of difficulty.

We used two criterion tests, one administered immediately after instruction, the other approximately two weeks later. Both the number of errors on the criterion tests and the time required to take the criterion tests were recorded. On the immediate criterion test, none of the main effects approached significance, but a significant interaction between amount of typographical cuing and difficulty level of the program did occur at the 5 percent level of significance. The addition of typographical cues to the Hard program improved criterion test performance, the addition of typographical cues to the Medium program reduced criterion test performance. Typographical cues on the Easy program made no difference on the criterion scores. We also discovered a significant linear trend in the amount of time taken to complete the criterion test. If response latency is any indication of the degree of learning, it would appear that taking a program constructed with complications in wording and terminology enables one to perform on the criterion test more rapidly than taking easier forms of the same program. However, by the time the second criterion test was administered, all significant differences in both criterion test score and in time required to complete the criterion test disappeared. In summary

then, we found that typographical cues reduce error rate on the program but improve cirterion tests performance only with a difficult program. They interfere with criterion test performance on a medium difficulty level program.

The third study was an attempt to define more precisely the specific variations in the program which cause it to be difficult. Two methods of making the program more difficult were subjected to experimental manipulation and test. The first method consisted of adding to each frame in the program one item of irrelevant information, producing what we may call our Irrelevant program. An item of irrelevant information was defined as some statement which was related to the topic under consideration but which was not necessary in order to solve the problem or answer the question in that frame. The second method, resulting in our Complex programs, consisted of requiring the student either to go through an additional step in order to solve the problem or to read deliberately complicated wording in the frames. The Control programs were written in simple straightforward language without any unnecessary complications or irrelevant information.

In order to obtain replication of results, three different content areas within educational measurement were programed by each of the three methods. Thus, nine different sections were prepared--three different content areas written in each of three different manners, Irrelevant, Complex, and Control. Then by combining three sections at a time in various combinations in accordance with a Latin-square design, nine different forms of the material were assembled. Each student received a booklet which contained all three content areas and all three methods

of presentation, but the particular combination and order of method and content for a particular student was assigned to him according to an elegant Latin-square which corresponded to Lindquist's Type V mixed design. Analysis of the error rate on the program revealed that we had indeed prepared programs of different difficulty levels in all three content areas. On the average, the Control programs had an error rate of about 8 percent; Irrelevant programs had an error rate of about 18 percent; and the Complex programs had an error rate of about 24 percent.

The criterion test was administered two days after completing the program. The criterion test consisted of nine 5-item sub-tests. Each sub-test covered one of the three content areas and was constructed by one of the same three methods of writing, Irrelevant, Complex, or Control. The resulting four-dimensional analysis of variance revealed no significant differences on the criterion test attributable to the different treatments. In summary then, this study failed to reveal significant differences in learning in spite of the fact that reliable differences in difficulty level had been produced by either the adding of irrelevant information or the adding of complex working and complex problems.

The fourth study was an attempt to determine whether variations in multiple-choice type programs might produce different results than constructed response programs. Within the multiple-choice programs we constructed variations along three dimensions: (1) the plausibility or implausibility of the alternatives, (2) the number of alternatives provided--either two or four, and (3) the form of the answer--either recording the letter of the correct alternative or writing out the

correct alternative in its entirety.

We found large consistent differences in the error rate on the program between the plausible and implausible alternative programs. As you would expect, the program was more difficult when students were asked to discriminate between more plausible alternative (14 percent). When the alternatives were less plausible, students made few errors on the program (4 percent). The number of alternatives also made a slight difference in error rate with four-choice alternative frames proving slightly more difficult (10 percent) than two-choice alternatives (8 percent). The form of the answering, whether by letter or by writing out the entire answer, did not produce significant differences in difficulty level.

Two types of criterion tests were prepared. One was a multiple-choice criterion test and the other was a constructed response criterion test. Alternate forms of each were prepared, one administered immediately after learning, the other two weeks later. The results of the criterion tests as based upon a preliminary analysis were not highly significant. On the immediate multiple-choice criterion test, writing out the answer proved to be more effective than writing merely the letter of the correct alternative ($p < .05$). Four alternatives proved better than two alternatives ($p < .10$). There was a tendency for the plausible program to produce better responses than the implausible but this did not reach conventional levels of significance. On the constructed response immediate test, none of the differences approached conventional significance levels. On the delayed multiple-choice test two weeks later, the plausible alternatives program appeared better

than the implausible ($p < .20$). On the delayed constructed response program, the plausible program appeared better than the implausible ($p < .15$). All other differences were non-significant.

While none of the differences in this study proved to be very large, it was interesting to note a consistent thread throughout all of them. The condition that produced the higher error rate also tended to produce better performance on the criterion tests. Thus, for example, providing more plausible alternatives on the program caused students to make more errors on the program but tended to improve their performance on the delayed constructed response criterion test. Requiring students to discriminate among four alternatives rather than two produced a higher error rate but increased their performance on the immediate multiple-choice criterion test. Asking students to write out a complete answer rather than merely writing the letter of the correct answer caused students to perform better on the multiple-choice immediate criterion test even though it did not affect their error rate on the program.

Although the results of all these experiments will need to be replicated by others before we can assign much confidence to them, it is interesting to summarize the trends that have been revealed so far. We can construct programs in many ways in order to vary their difficulty level. Some of these ways increase criterion performance, some decrease criterion performance, and for some we have been unable to demonstrate any effect in either direction.

1. What factors affect difficulty level without appearing to affect criterion performance? Adding irrelevant information and

requiring more complex reading and problem solving increases error rate, but we have not found a significant effect on criterion performance. Providing additional clues and hints decreases error rate but has not been found to affect criterion performance.

● 2. What factors increase the error rate and also increase criterion performance? Increasing both the error rate and criterion performance may be done by asking the student to discriminate among plausible rather than implausible alternatives and by increasing the number of alternatives from two to four.

3. What decreases the error rate and decreases learning also? We found that asking students to make trivial responses on the program reduces the error rate but significantly reduces test performance also. We found that typographical cues also decrease both error rate and test scores when applied to a medium-level difficulty program; when typographical cues are applied to a hard program, they still decrease difficulty level but increase criterion performance.

It is clear that the difficulty level of a program is no criterion for assessing the amount of learning the program will produce, but the manner of writing the program is important.

Chapter VII

Evaluation of Programmed Instruction

by

John D. Krumboltz

Originally published in J. P. Lysaught (Ed.) Programmed Instruction in Medical Education. Rochester, N. Y.: Rochester Clearing House, University of Rochester, 1965, Pp. 139-149.

Chapter VIII

Needed Research in Programed Instruction

by

John D. Krumboltz

Originally published in Educational Leadership, 1963, Vol 21., No.1, 30-33, 49.

Chapter IX

The Comparative Effects of Inductive And Deductive Sequences in Programed Instruction

by

John D. Krumboltz and William W. Yabroff

Originally published in the American Educational Research Journal, 1965, Vol. 2, No. 4, 223-235.

Appendix A

Inductive Mixed Arrangement of

Programed Booklet

IND-M

Name of student _____

P R O G R A M E D B O O K L E T

Interpreting Test Results Part 1

Form IND-M

John D. Krumboltz

William W. Yabroff

-Stanford University-

Note: Read Instructions Carefully

DO NOT WRITE IN THIS BOOKLET

How to Use this Programed Booklet

You are about to begin studying by a new method. Programed books are not arranged like ordinary books. To use this booklet, follow these steps carefully:

1. Turn to page one. Read the words in the top box (Frame 1) only. Think of some response which completes the blank space in Frame 1.

   Write down that response on your answer sheet.

2. Turn to page 2. Do not read any more on page 1 now. In the top left box on page 2, you will find the correct answer to Frame 1.

3. If your response agrees with the correct answer, write a plus sign (+) next to your response on the answer sheet. If your response does not agree, write a minus sign (-).

4. Now read Frame 2 at the top of page 2. Write down a response to Frame 2 on your answer sheet and score it + or - according to the correct answer given at the top left of page 3.

5. Continue in this manner to the last page of the booklet. The answer to this frame is found back on page 1 in the next to the top square on the left hand side.

   Be sure you always write down your answer before looking at the answer in the booklet.

After completing all the frames in the booklet, return the booklet and the answer sheet to your instructor.

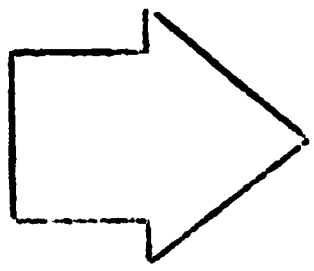
DO NOT WRITE IN THE PROGRAMED BOOKLET

Important

The following symbols tell you what kind of a response is needed:

1. The number of words needed to complete an item is indicated by the number of blanks. Thus, _____ indicates a one-word response, whereas _____ _____ indicates a two-word response. When you see three asterisks (* * *), you are to use as many words as you think necessary to respond to the item.
2. The number sign (#) indicates that the desired response is a number.
3. The abbreviation (TT) calls for a technical term. When it is used, a nontechnical word is wrong.
4. Often several responses are equally good, even though all of them are not listed in the answer box. This is particularly true when the response is nontechnical. Use reasonable judgment in deciding whether your response means the same as the correct answer. Score it correct if it does.

DO NOT WRITE IN THE PROGRAMED BOOKLET.

	On the 5th grade spelling test, George spelled 20 words correctly. George's raw score is 20. Mary spelled 27 words correctly. Mary's raw score is _____(#). DO NOT READ ANY MORE ON THIS PAGE NOW. TURN TO THE TOP OF PAGE 2 TO CONFIRM YOUR ANSWER. 1
(1) items (questions) (2) group CONTINUE ALONG SECOND ROW---- 24	By giving the percent of the scores that fall below a given score, a percentile score tells the relative standing of a (1)_____ with some (2)_____. 25
(1) 6 (2) the same 48	Suppose that Philbrite obtained an unusual score of 27 rather than the usual score of 11 for a bright 3rd grader, so that scores for the 3rd grade class were: Nancy 1 Jack 5 Roger 7 Philbrite 27 The mean and median scores are (the same, different)_____. 49
percentile 72	To avoid the misunderstandings which arise from the misuse of grade equivalent scores, (1)_____ (TT) scores (properly used) would compare the performance of an (2)_____ with the performance of some (3)_____ to which he belongs or aspires to belong. 73
26 96	Frank obtained a raw score of 11 on the Delta reading test. Frank's percentile score is (1)_____(#). How many 7th graders in the standardization sample scored less than Frank? (2)_____(#). 77 97

ANSWER 27

NOW READ FRAME 2 RIGHT HERE-->

1

A raw score of 18 means that an individual obtained _____(#) correct answers on a test.

NOW TURN TO THE TOP OF PAGE 3 TO CHECK YOUR ANSWER.

2

(1) person

(2) group

25

Trudy took the College Entrance Examination at Little Run College and at Honors College. She obtained the following scores:

Little Run College	Honors College
percentage score = 85	percentage score = 90
percentile score = 88	percentile score = 75

At which college did Trudy score highest among the other applicants who took the Entrance Examination? _____.

26

different

49

Two types of average scores are (1)_____(TT) and _____(TT) scores which have about the same value in a (2)_____ distribution. In an unusual distribution of scores, the mean is (3) * * *)

50

(1) percentile

(2) individual

(3) group

73

Suppose you wished to know how well 5th graders perform on the Sanford Achievement Test. One way would be to give this test to the entire 5th grade population and compute an average score. However, it is seldom if ever possible to test the entire _____(TT) of 5th graders.

74

(1) 40

(2) 400

97

A 7th grade boy would have to obtain a raw score of _____(#) to score at the median of the standardization sample.

98

18

CONTINUE READING THE TOP FRAMES ON EACH PAGE - - →

2

The number of correct answers obtained on a test is known as the _____(TT) score.

NEVER READ TWO FRAMES CONSECUTIVELY ON THE SAME PAGE

3

Little Run College

26

Honors College admits 570 new students a year, and rejects 64 percent of their applicants on the College Entrance Examination. The lowest acceptable percentile score earning admission would be _____(#).

27

(1) mean, median

(2) usual, (or typical)

(3) more affected by extreme scores(or words to that effect)

50

Below are the results of Miss Reallybusy's findings for grades 3, 5, and 7 on the Jones Arithmetic Test:

Grade 3	Grade 5	Grade 7
Mean = 6	Mean = 11	Mean = 16
Median = 6	Median = 10.5	Median = 16.5

Both the mean and median scores for each grade report two kinds of _____.

51

population

74

It is possible to give the Sanford Achievement test to different 5th grade classes, or samples of 5th graders. By taking different (1)_____(TT) we would be able to describe how well 5th graders as a whole, or technically the 5th grade (2)_____(TT) might be expected to perform on the test.

75

13

98

Thus, norms allow us to compare an (1)_____ score to the scores of the (2)_____ _____(TT) from which the norms were computed.

99

CONTINUE------>

raw

3

Frank obtained a raw score of 20 on a 50-item English test. His percentage score is obtained as follows:

$$\frac{20}{50} \times 100 = 40$$

Mary obtained a raw score of 10 on the same test. Her percentage score is _____(#).

4

6"

27

A score that gives the percent of the scores falling below a given score is called a (1)_____(TT) score, which tells the relative standing of a person with some (2)_____.

28

---

averages

51

If Philbrite in the 3rd grade had obtained an unusual score of 27 (very high for a 3rd grader) Miss Reallybusy would have found the following:

Grade 3	Grade 5	Grade 7
Mean = 10	Mean = 11	Mean = 16
Median = 6	Median = 10.5	Median =16.5

The mean score for the 3rd grade is now very close to the mean score of the (1)_____grade. The (2)_____ score for the 3rd grade has not changed in spite of Philbrite's performance.

52

---

(1) samples

(2) population

75

In this case, the various 5th grade classes or the _____(TT) of 5th graders are selected from all possible 5th graders, to find out how well 5th graders in general might be expected to perform on the Sanford Achievement Test.

76

---

(1) individual's

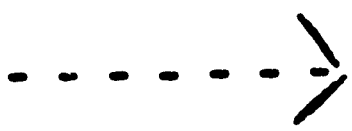
(2) standardization sample

99

Now turn to Panel #3, page 27. Panel #3 reports (1)_____(TT) norms for a standardization sample of (2)_____(#) 7th grade girls on the Delta Reading Test.

100

20 (10/50 x 100 = 20)



4

Larry Bright missed two answers on a 75-item quiz. Larry's raw score is (1)_____(#). His percentage score is obtained by dividing 75 by (2)_____(#) then multiplying by 100.

ALWAYS TURN TO THE NEXT PAGE

5

(1) percentile

(2) group

28

In the following test scores:

12
13
15
19
24

the score at the median is 15. There are _____(#) scores above and below 15.

29

(1) 5th

(2) median

52

If Miss Reallybusy reported that the 3rd grade performed as well as the 5th grade on the Jones Arithmetic test, what kind of average score would she have used in this report? _____(TT)

53

samples

76

Thus, a (1)_____(TT) is used to describe or obtain information about a (2)_____(TT) and is comprised of subjects selected from that (3)_____(TT).

77

(1) percentile

(2) 800

100

A girl would have to obtain a score of at least _____(#) to fall at the 95th percentile.

101

(1) 73 (2) 73 AND SO ON-----> 5	A percentage score is the (1)________(TT) of correct answers on a test, whereas a raw score is the (2)________of correct answers. YOU WILL NEVER GET LOST IF YOU FOLLOW THESE FRAME NUMBERS- - - - - - - - - - - - - - - - - -> 6
2 29	In scores 7 10 11 14 19 the score at the median is ______(#). 30
mean 53	In a typical distribution of test scores, the (1)__________(TT) and __________(TT) have about the same value. The (2)__________(TT) is affected more by an occasional extreme score. 54
(1) sample (2) population (3) population 77	A foreign observer found that high school students in Southville County averaged one hour of homework each night. He was asked to make a statement about the amount of homework done by high school students in the United States. In this case, the sample used was high school students in (1)__________. The population to be described was high school students in (2)_______________. 78
29 101	Mary obtained a raw score of 20 on the Delta Reading test. How many girls in the standardization sample scored below Mary? _________(#). 102

(1) percent
(2) number

Susan answered 16 out of 20 test-items correctly. Her percentage score is _____(#).

CORRESPONDING ANSWER NUMBER → 6

7

11

When scores are not arranged in order, such as 3, 9, 8, 4, 11, the median is found by ranking the scores from lowest to highest as follows:

Scores	Rank Order
3	11
9	9
8	8
4	4
11	3

30 The median is _____(#). 31

(1) mean, median
(2) mean

Books read per month in Southville School

7th grade	8th grade
2 students read 25	3 students read 3
13 students read 1	14 students read 1
5 students read 0	3 students read 0

What kind of average score would report both classes to have read an equal number of books per month? (1)_____(TT). This score indicates that each class read (2)_____(#) books per month.

54 55

(1) Southville County
(2) the United States

Before making a statement about the amount of homework done by high school students in the United States, it would be wise for the observer to have many _____(TT) of high school students from different parts of the country.

78 79

600

Norms allow us to compare an individual's score on a test to the scores of the _____ _____(TT) from which the norms are computed.

102 103

80 7	In an 80 question History exam, Frank obtained a percentage score of 70. His raw score was _____(#). If you have read this far, circle number 8 on your answer sheet like this: (8). 8
8 31	A score is said to be at the median when (* * *) 32
(1) median (2) 1 55	The Southville Daily Times wrote a stirring editorial claiming that the 7th grade class in Southville school read almost three times as many books per month as the 8th grade class. Using the same figures as on the previous frame, what kind of average score did the newspaper use to make this observation? (1)_____(TT). This score indicated that the 7th grade read on the average of (2)_____ books per month while the 8th grade averaged (3)_____ books per month. 56
samples 79	It is seldom possible to obtain information by testing an entire population. It is possible to use (1)_____(TT) to describe certain characteristics of a (2)_____(TT). 80
standardization sample 103	Refer again to Panels #2 and #3. In comparing raw scores needed to obtain a percentile score of 95 on the Delta Reading Test, would a boy or a girl have to obtain a higher raw score to be at the 95th percentile? _____. 84

56

To obtain a percentage score, divide the (1)_____(TT) score by the (2)_____ number of test items, then multiply by (3)_____(#).

8 | 9

there are an equal number of scores above and below that score. (or words to this effect)

Arrange the following even number of scores in rank order:

Scores	Rank Order
19	_____
27	_____
12	_____
15	_____

What number would represent a median score? _____(#)

32 | 33

(1) mean

(2) 3 or 3.15

(3) 1 or 1.15

The mean and median are two kinds of averages which may be the same (or close together) in a usual set of scores, but may be _____ in an unusual set of scores.

56 | 57

(1) samples

(2) population

He might find that students in Northville average 3 hours of homework per night, while students in Westville averaged 2.5 hours etc. It is obvious that these samples are _____.

If you have read this far, circle number 81 on your answer sheet.

80 | 81

girl

Frank and Susan both obtained raw scores of 21 on the Delta Reading Test. When compared to their standardization samples, who obtained the higher percentile score? _____

85

104 | 105

(1) raw (2) total (3) 100 9	Frank's percentage score of 70 on the 80-item History Quiz enables us to compute the (1)_____ of correct answers he obtained on the test. Do we know how well Frank did in comparison to other members of his class? (2)_____. 10
(1) 27, 19, 15, 12 (2) 17 33	In the same set of scores: 12 15 19 27 the median of 17 is half-way between the two middle-most scores, 15 and 19. The median of 17 is a midpoint at which _____(#) scores fall above and below the median. 34
different 57	Turn to Panel #1 on page 25. Which 7th grade student obtained a raw score equivalent to the mean of the 7th grade? (1)_____. Thus, a raw score of (2)_____ would have a grade equivalent of 7, because it is equal to the (3)_____(TT) score of the 7th grade. 58
different 81	Since it would be practically impossible to know the homework time spent by every high school student in the United States, the sample of students in Southville might be compared with other _____(TT) of students throughout the country. 82
Susan 105	If you were assigning an A to boys and girls at or above the 90th percentile on the Delta Readint Test, what minimum raw score would a girl have to obtain? (1)_____(#) A boy? (2)_____(#) 106

(1) number (2) no 10	In the 4th grade class of 30 pupils, Trudy answers 15 out of 30 test-items correctly. Trudy's percentage score is (1)_____(#). Do we know how well Trudy did in comparison to the other 30 pupils? (2)_____. 11
2 34	Arrange the following even number of scores in rank order and find the median: Scores: 9, 3, 13, 47 Rank Order: 47, ____, ____, ____ The median is (1)_____(#). The two middle-most scores are (2)_____ and _____(#). 35
(1) James (2) 30 (3) mean 58	A student with a raw score of 15 on the Sanford Achievement test did as well as the average of the (1)_____(#) grade, and has a grade equivalent of 3, because 15 is the (2)_____(TT) score made by 3rd graders. 59
samples 82	Thus, a group of subjects selected from a population for the purpose of obtaining information or describing that population is known as a _____(TT). 83
(1) 25 (2) 24 106	Thus, when standardization samples (1)_____ significantly from each other, more than one set of (2)_____(TT) should be given for a test. 87 107

(1) 50
(2) no

11

The percent of correct answers obtained on a test, found by $\frac{\text{raw score}}{\text{total score}} \times 100$ is the _____(TT) score.

12

Rank Order
47
13
9
3

(1) 11
(2) 9, 13

35

Thus, the median in an even number of scores is (1)_____ between the two (2)_____ scores.

36

(1) 3rd
(2) mean

59

A student who obtained a raw score of 20 did as well as the average of the (1)_____(#) grade, and has a grade equivalent of (2)_____(#).

60

sample

83

Miss Reallybusy reported that her 3rd grade class norm on the Sanford Achievement Test was a mean score of 45. The standardization sample from which the class norm of 45 was computed is the (1)_____ _____.

84

(1) differ
(2) norms

107

It is useful to have separate (1)_____(TT) for boys and girls on the Delta Reading Test because the two sexes make significantly (2)_____ scores. If you have read this far, circle number 108 on your answer sheet.

108

percentage 12	On the World Geography Test, 15 out of 20 students (75 percent of the class) scored lower than George. His percentile score is 75. Ten students scored lower than Susan. What percent of the class scored below Susan? (1)_____(#). Her percentile score is (2)_____(#). 13
(1) halfway (2) middlemost 36	In scores 3, 4, 4, 4, 5, 7, 8, 11, 11, 12, applying the definition you have already learned about the median, would you count the duplicate scores 4 4 4 and 11 11 as separate scores? (1)_____. At what percentile would the median fall? (2)_____(#). The median for the above set of scores is (3)_____(#). 37
(1) 5th (2) 5 60	A grade equivalent indicates (* * *) 61
3rd grade 84	The superintendent of Southville School District reported that the Southville High School's senior class scored at the 83rd percentile on the District Norm for the Sanford Achievement Test. Senior classes in the Southville School District were used as the _____ _____(TT) for the district norm. 85
(1) norms or percentile norms (2) different 108	Might it be useful to have more than one set of norms for the Jones Arithmetic Test when comparing scores obtained by Miss Reallybusy's students in Northville and scores obtained by students in Far-behind School at Reservationville? _____. 109

(1) 50 (2) 50 13	In the same class of 20 students, Frank obtained a percentile score of 90 on the World Geography Test. What percent of the class scored lower than Frank? _____(#). 14
(1) yes (2) 50th (3) 6 37	In the following set of scores: 7 19 7 100 8 100 15 100 The median is (1)_____(#). The 50th percentile is (2)_____(#). 38
which grade obtained a mean score closest to an individual's score (or words to this effect) 61	Since Miss Reallybusy administered the Sanford Achievement test at the beginning of the school year, a raw score of 15 is equivalent to the mean of beginning 3rd graders. Thus, the _____ _____(TT) may be more precisely expressed as 3.0. 62
standardization sample 85	The test publishers of the Sanford Achievement test report a table of national percentile norms for high school seniors. What standardization samples were probably used to compute these norms? (* * *) 86
yes 109	More than one set of norms should be given for a test when the standardization samples _____ significantly from each other. 90 110

90 14	Thus, you might say that a percentile score is the percent of (* * *). 15
(1) 17 (2) 17 38	In Frank's class of 24 students, Frank scored 18 on the Know Your Geography Quiz. 12 students scored less than Frank. The median score for Frank's class is (1)_____(#). At what percentile did Frank score? (2)_____(#). 39
grade equivalent 62	If Miss Reallybusy had administered the test after two months of the school year had passed and found that 15 was the mean raw score made by 3rd graders, then 3.2 would be the _____ _____(TT). 63
senior classes throughout the nation (or words to that effect) 86	The sample of subjects on whose scores the (1)_____(TT) of a test are computed is known as a (2) _____ _____(TT). 87
differ 110	Susie obtained a raw score of 58 on the College Entrance Test. When she applied for admission to Outeast, Somefun and Honors colleges, she learned that her raw score of 58 earned a different percentile score at each college. Susie's raw score was compared with at least (1)_____(#) different sets of (2)_____(TT). 111

the scores that fall below a given score. (or words to this effect) 15	Turn to Panel #2 on page 26. The raw score of 13 would have a percentile score of _____(#) on the Delta Reading Test. 16
(1) 18 (2) 50th 39	A score is said to be at the median when there are an (1)_____ number of scores above and below that score. The median is at the (2)_____(#) percentile. 40
grade equivalent 63	If a raw score of 32 had a grade equivalent of 7.6, we would know that 32 was the (1)____(TT) score obtained by the (2)_____ grade after (3)____(#) months of school had passed. 64
(1) norms (2) standardization sample 87	Miss Reallybusy's 3rd grade class norm on the Sanford Achievement test was a mean score of 45. The class norm computed from the scores of the 3rd grade standardization sample is the (1)_____(TT) score of (2)_____(#). 88
(1) 3 (2) norms 111	Susie's percentile scores for the College Entrance Test were as follows: Outeast College, 64; Somefun College, 72; Honors College, 45. At which college would Susie compare most favorably to other applicants for admission? (1)_____. Least favorably? (2)_____. 92 112

50 16	On Panel #2, what percent of the scores are below the 60th percentile? (1)_____(#). What percent of the scores are below the raw score of 10? (2)_____(#). 17
(1) equal (2) 50th 40	The <u>mean</u> of scores 2, 4, and 9 is obtained as follows: $\frac{2+4+9}{3}$ = _____(#) 41
(1) mean (2) 7th (3) 6	A grade equivalent indicates which (1)_____ obtained a (2) _____(TT) score closest to an individual score.
64	65
(1) mean (2) 45 88	The percentile score of 83 obtained by the senior class of Southville High was found by comparing the scores of the senior class to other senior classes in the Southville High School District. The percentile scores of these standardization samples constitute a set of district _____(TT) for the Sanford Achievement Test. 89
(1) Somefun (2) Honors College 112	Thus, a test which provides many sets of (1)_____(TT) based on (2)_____ standardization samples allows more useful (3)_____ of a person's score. 93 113

(1) 60 (2) 35 17	Thus, a percentile score is the (1)_____(TT) of the scores that fall (2)_____ a given score. 18
5 41	The mean for scores 3 4 6 7 10 is obtained by dividing 30 by (1)_____(#). The mean is (2)_____(#). 42
(1) grade (2) mean 65	On Panel #1, page 25, Fidel (5th grade) obtained a raw score of 15. His grade equivalent is (1)_____(#), because he did as well as the average of the (2)_____ grade. Would you say on the basis of his grade equivalent score, that Fidel is two years retarded? (3)_____. 66
norms 89	In reporting percentile scores obtained by senior classes in high schools through the Nation, the test publishers of the Sanford Achievement test were reporting a set of national _____ _____(TT) for that test. 90
(1) norms (2) different (3) comparison (or evaluations) 113	The Know Your Literature Test reported a set of percentile norms for graduate students at Oxford, England. Miss Earnest administered the test as a final examination to the Junior Literature class at Southville High and found that 95 percent of the students scored below the 10th percentile! "A whole year of teaching wasted," she concluded. This test (did, did not)_____ provide useful comparisons of scores. 114

(1) percent (2) below 18	In a 40-item English test, Frank obtained a percentage score of 90. Do we know Frank's relative standing with the other members of his class? _____. 19
(1) 5 (2) 6 42	The mean is obtained by (* * *). 43
(1) 3.0 (2) 3rd (3) no 66	Since grade equivalents are based on mean scores, we can expect that approximately _____(#) percent of a class will fall below the class mean. Should these students be compared in their performance to the performances of students in lower grades? 67
percentile norms 90	Thus, the scores obtained by standardization samples on a test constitute the _____(TT) of the test. 91
did not 114	A reporter for the Southville Daily Times wrote that the performance of the Junior Class on the much acclaimed "Know Your Literature" test proved that education in Southville High was only one-tenth as good as English Education! The lack of appropriate _____(TT) for the "Know Your Literature" test led this reporter to a misinterpretation of test scores. 115

no 19	If we knew Frank's percentile score on the 40-item English test to be 80, do we now know Frank's relative standing with his class? (1)_____. What percent of the class scored below Frank? (2)_____(#). 20
dividing the sum of the scores by the number of scores (or words to this effect) 43	In Frank's class of 20 students, 12 obtained raw scores of 10, and 8 students obtained raw scores of 5 on the Friday Vocabulary Quiz. Frank decided to compute the mean for his class as follows: $\frac{10+5}{20} = .75$. Is this the correct mean for Frank's class? _____. 44
50 67	Imagine being told that you could solve a puzzle as fast as the average monkey. You might well reply, "That would be valuable information if I were a monkey, but how did I rate in relation to _____ of my own age, sex, and educational level? 68
norms 91	Turn now to Panel #2, page 26. What standardization sample was used for the Delta Reading Test? ___ _____ _____. 92
norms 115	Miss Earnest should have chosen a literature test which provided many sets of (1)_____(TT) based on different (2) _____ _____(TT) other than Oxford Graduate Students. 116

(1) yes
(2) 80

20

By giving the percent of the scores that fall below a given score, a percentile score tells (* * *)

21

no

44

In Computing the mean of .75 on the previous frame, Frank failed to find the _____ of the scores. The correct mean for Frank's class on the Friday Vocabulary Quiz is 8.

45

persons (humans)

68

Grade equivalents are widely used and misused in education. The trouble with grade equivalent scores is that they may compare the performance of an (1)_____ _____ to the performance of some (2)_____ to which he (does, does not) (3)_____ belong.

69

7th grade boys

92

On Panel #2, the scores obtained by 7th grade boys on the Delta Reading Test constitute a set of _____ _____(TT) for that test.

93

(1) norms
(2) standardization samples

116

A test which provides many sets of norms based on _____ standardization samples allows for more useful comparisons of a person's score.

97

117

the relative standing of a person with some group (or words to this effect) 21	Would it be possible for a person to answer 95 percent of the items correctly on a test and only obtain a percentile score of 20? _____. 22
sum 45	The mean is the (1)_____ of the scores divided by the (2)_____ of scores. If you have read this far,circle number 46 on your answer sheet. 46
(1) individual (2) group (3) does not 69	In Miss Reallybusy's 7th grade class (Panel #1) Hilda's raw score of 24 might be reported either as a grade equivalent of 5.8 or as a percentile score of _____(#). 70
percentile norms 93	The letter _____ in Panel #2 points to the number of subjects in the standardization sample. (Hint: Remember, there were 1,000 in the sample) 94
different 117	Congratulations! You have just finished your programmed instruction booklet. Please answer the questions found on the last page of your answer sheet. 98 Thank you. 118

yes 22	Again, would it be possible for a person to answer only 30 percent of the items correctly on a test and obtain a percentile score of 85? _____. 23
(1) sum (2) number 46	Miss Reallybusy teaches grades 3, 5, and 7. Below are raw scores for grade 3 on the Jones Arithmetic test: Nancy 1 Roger 7 Jack 5 Philbrite 11 To find the average score for the 3rd grade, Miss Reallybusy computed the median score. The 3rd grade median is (1)_____(#). She might also find another kind of average by computing a (2)_____(TT) score. 47
20 70	In reporting Hilda's raw score of 24 as a percentile score of 20, we are comparing her performance to the performance of the 7th grade, a (1)_____ to which she (2)_____. 71
N 94	Thus, the (1)_____(TT) of a test are computed from scores obtained by the (2)_____ _____(TT) on that test. 95
  118	

yes 23	The percentage score on a test tells us the percent of the (1)_____ answered correctly. The percentile score tells us the relative standing of an individual with some (2)_____. ANSWER IS IN 2ND ROW DOWN, PAGE 1. TURN TO PAGE 1 AND BEGIN 2ND ROW. 24
(1) 6 (2) mean 47	The mean score for the 3rd grade is: Nancy 1 Jack 5 Roger 7 Philbrite 11 Mean = (1)_____(#). The mean and median scores for the 3rd grade are (the same, different). (2)_____. 48
(1) group (2) belongs 71	Jim, although still a high school senior, scored at the 85th percentile of freshman applicants to Honors College. The use of a _____(TT) score to report Jim's performance on the College Entrance Test compared him with a group to which he aspired to belong. 72
(1) norms (2) standardization sample 95	On Panel #2, a 7th grader must have a raw score of at least _____(#) to obtain a percentile score of 95. 96

Appendix B

Panels Accompanying Programed Booklets

P A N E L #1

Raw Scores on the Sanford Achievement Test

Grade 3

Name	Raw Score
Nancy	10
Jack	13
Roger	15
Philbrite	22
Total	60

3rd grade mean = 15

Grade 5

Fidel	15
Mary	18
John	20
Louise	22
Gene	25
Total	100

5th grade mean = 20

Grade 7

Gary	16
Hilda	24
James	30
Janice	38
Don	42
Total	150

7th grade mean = 30

Percentile Norms for 7th Grade Boys on the Delta Reading Test (N = 1000)

Percentile Score	Raw Score
99	33 or higher
97	29 - 32
95	26 - 28
90	24 - 25
85	23
80	22
75	21
70	19 - 20
65	17 - 18
60	15 - 16
55	14
50	13
45	12
40	11
35	10
30	9
25	8
20	7
15	6
10	5
5	4
3	3
1	0 - 2

Percentile Norms for 7th Grade Girls on the Delta Reading Test (N = 800)

Percentile Score	Raw Score
99	37 or higher
97	34 - 36
95	29 - 33
90	25 - 28
85	23 - 24
80	21 - 22
75	20
70	19
65	18
60	17
55	16
50	15
45	14
40	13
35	12
30	11
25	10
20	8 - 9
15	7
10	6
5	5
3	3 - 4
1	0 - 2

Appendix C

Answer Sheet

IND-M

A N S W E R S H E E T

Interpreting Test Results Part 1

Form IND-M

Keep a record here of the time you spend on this program.

Date	Time at Start	Time at Finish	Number of Minutes spent
_____	_____	_____	_____
_____	_____	_____	_____
_____	_____	_____	_____
_____	_____	_____	_____

Total time _____

Name _____

Year in college (circle one) 1 2 3 4

MAS _____
(leave blank)

	+ or -
1. _____	( )
2. _____	( )
3. _____	( )
4. _____	( )
5. 1) _____	( )
2) _____	( )
6. 1) _____	( )
2) _____	( )
7. _____	( )
8. _____	( )
9. 1) _____	( )
2) _____	( )
3) _____	( )
10. 1) _____	( )
2) _____	( )
11. 1) _____	( )
2) _____	( )
12. _____	( )
13. 1) _____	( )
2) _____	( )
14. _____	( )
15. _____	
_____	
_____	( )
16. _____	( )

	+ or -
17. 1) _____	( )
2) _____	( )
18. 1) _____	( )
2) _____	( )
19. _____	( )
20. 1) _____	( )
2) _____	( )
21. _____	
_____	
_____	( )
22. _____	( )
23. _____	( )
24. 1) _____	( )
2) _____	( )
25. 1) _____	( )
2) _____	( )
26. _____	( )
27. _____	( )
28. 1) _____	( )
2) _____	( )
29. _____	( )
30. _____	( )
31. _____	( )
32. _____	
_____	
_____	( )

33. (Rank Order)___ ___ ___ ___ ( ) + or -

_____ ( )

34. _____ ( )

35. (Rank Order)___ ___ ___ ___ ( )

1) _____ ( )

2) _____ ( )

36. 1) _____ ( )

2) _____ ( )

37. 1) _____ ( )

2) _____ ( )

3) _____ ( )

38. 1) _____ ( )

2) _____ ( )

39. 1) _____ ( )

2) _____ ( )

40. 1) _____ ( )

2) _____ ( )

41. _____ ( )

42. 1) _____ ( )

2) _____ ( )

43. _____

_____

_____ ( )

44. _____ ( )

45. _____ ( )

46. _____ ( ) + or -

47. _____ ( )

48. 1) _____ ( )

2) _____ ( )

49. _____ ( )

50. 1) _____ ( )

2) _____ ( )

3) _____

_____ ( )

51. _____ ( )

52. 1) _____ ( )

2) _____ ( )

53. _____ ( )

54. 1) _____ ( )

2) _____ ( )

55. 1) _____ ( )

2) _____ ( )

56. 1) _____ ( )

2) _____ ( )

3) _____ ( )

57. _____ ( )

58. 1) _____ ( )

2) _____ ( )

3) _____ ( )

59. 1) _____ ( )

2) _____ ( )

+ or -

60. 1) _____ ( )

2) _____ ( )

61. _____

_____

_____ ( )

62. _____ ( )

63. _____ ( )

64. 1) _____ ( )

2) _____ ( )

3) _____ ( )

65. 1) _____ ( )

2) _____ ( )

66. 1) _____ ( )

2) _____ ( )

3) _____ ( )

67. _____ ( )

68. _____ ( )

69. 1) _____ ( )

2) _____ ( )

3) _____ ( )

70. _____ ( )

71. 1) _____ ( )

2) _____ ( )

72. _____ ( )

73. 1) _____ ( )

2) _____ ( )

3) _____ ( )

+ or -

74. _____ ( )

75. 1) _____ ( )

2) _____ ( )

76. _____ ( )

77. 1) _____ ( )

2) _____ ( )

3) _____ ( )

78. 1) _____ ( )

2) _____ ( )

79. _____ ( )

80. 1) _____ ( )

2) _____ ( )

81. _____ ( )

82. _____ ( )

83. _____ ( )

84. _____ ( )

85. _____ ( )

86. _____

_____

_____ ( )

87. 1) _____ ( )

2) _____ ( )

88. 1) _____ ( )

2) _____ ( )

89. _____ ( )

90. _____ ( )

91. _____ ( )

	+ or -
92. _____	( )
93. _____	( )
94. _____	( )
95. 1) _____	( )
2) _____	( )
96. _____	( )
97. 1) _____	( )
2) _____	( )
98. _____	( )
99. 1) _____	( )
2) _____	( )
100. 1) _____	( )
2) _____	( )
101. _____	( )
102. _____	( )
103. _____	( )
104. _____	( )
105. _____	( )
106. 1) _____	( )
2) _____	( )
107. 1) _____	( )
2) _____	( )
108. 1) _____	( )
2) _____	( )
109. _____	( )
110. _____	( )

	+ or -
111. 1) _____	( )
2) _____	( )
112. 1) _____	( )
2) _____	( )
113. 1) _____	( )
2) _____	( )
3) _____	( )
114. _____	( )
115. _____	( )
116. 1) _____	( )
2) _____	( )
117. _____	( )

PLEASE ANSWER THE QUESTIONS FOUND ON THE NEXT PAGE OF THIS ANSWER SHEET.

Appendix D

Evaluation Form

On the following questions, place an "X" next to the phrase which best describes your personal reaction to the programed booklet.

1. I feel the material contained in this booklet was

_____ a) very interesting.
_____ b) of some interest.
_____ c) of little or no interest.

2. If I become a teacher, I think that knowing how to interpret tests would be

_____ a) absolutely required.
_____ b) extremely important.
_____ c) of some importance.
_____ d) of little importance.

3. Think back to the sequence of definitions and problems as they appeared in your booklet. In learning the material, this sequence was

_____ a) very helpful.
_____ b) of some help.
_____ c) of little or no help.

4. The sequence of definitions and problems as they appeared in this program

_____ a) forced me to think most of the time.
_____ b) required some thought on my part.
_____ c) enabled me to get the correct answer without thinking very hard.
_____ d) required practically no thought on my part.

5. I would have learned the material better if

_____ a) I could have solved problems before learning the rule.
_____ b) I could have learned the rule before solving problems.
_____ c) (the order was satisfying as it was).

6. In comparing this booklet with the usual textbook way of learning new material, I felt I learned

_____ a) more efficiently from the booklet.
_____ b) as efficiently from the booklet as from a typical textbook.
_____ c) less efficiently from the booklet than from a typical textbook.

PLEASE RETURN YOUR BOOKLET AND ANSWER SHEET TO THE INSTRUCTOR.

Name of student _____
(please sign here)

112

Appendix E

Criterion Tests

C R I T E R I O N T E S T

F O R

I N T E R P R E T I N G T E S T R E S U L T S

John D. Krumboltz

William W. Yabroff

-Stanford University-

Name of student _____

Section R

Name of student _____

Exact time at beginning
of this section is _____

Answer the following questions in the spaces provided:

1. To compute the percentage score on a test, one must know the number of correct answers obtained and also the

   _____.

2. In a series of scores where the mean and median are obviously different, extreme scores affect the (1)_____ more than the (2) _____.

3. When information is sought on large numbers of subjects, and it is not feasible to examine each subject, investigators use what is known as a _____ to describe the population involved.

4. The percentage score and the raw score are two ways of reporting test results. The difference is that a percentage score tells

   (1) _____

   whereas a raw score gives (2)_____

   _____.

5. An individual's rank in some group is expressed by a _____ score.

6. When a student is assigned a grade equivalent score, his performance on a test is compared with a (1)_____ that obtained a (2) _____ score closest to his score.

7. A percentile score may be defined as _____

   _____.

8. If you had an even number of test scores, no two of which were identical, could the mean fall at one of these scores? (1)_____. Could the median fall at one of these scores? (2) _____.

9. The median is defined as _____

_____.

10. In comparing the raw score of a student with the mean score obtained by some class, grade equivalent scores may create a false impression. A student may score at a grade equivalent above or below his correct grade placement. Two factors about grade equivalent scores which account for this are:

(1) _____

_____ and

(2) _____

_____.

11. If you wished to compare an individual's score with a group to which he either belongs or aspires to belong, you would refer to norms reported in a table of _____ scores.

12. Which type of average score always corresponds to the 50th percentile? _____.

13. A mean score on a test is obtained by _____

_____.

14. The norms of a test are computed from the scores of subjects known as a _____ _____.

THE TIME AT THE COMPLETION OF THIS SECTION IS _____.

Section A

Name of student _____

Exact time at beginning of this section is _____

(Use space in the margin for computation if needed.)

Questions 1 through 5 are based on the table below which gives the names and raw scores of all 7th grade students who took the 50-item history examination at Midville Junior High.

Student	Raw Score
Adele	56
Roscoe	46
George	43
Betty	40
Katherine	33
Harry	31
Jeff	26
Irvin	21
Arthur	19
John	15
Sum of raw scores =	330

1. What was Betty's percentage score? _____

2. Give the percentile score for Arthur. _____

3. Give the percentile score for George. _____

4. To score at the median in this set of scores, one would have to obtain a raw score of _____.

5. For a student to have a grade equivalent score of 7.0, what raw score would he have to obtain if the test above was administered at the beginning of the school year? _____

6. Suppose you have written a test of 25 items for your class of 20 pupils, and you find that a raw score of 21 correct answers gave a percentile score of 30. Generally speaking, was this a hard or an easy test for your class? _____.

7. A journal reports that 19 out of every 20 teenage American boys score 50 or below on a physical fitness exam. If the score of 50 is a percentile score, the norms of the test (were, were not)

   (1)_____ computed from a representative sample of

   (2)_____.

8. To find a student's percentile score on the Tea Test of Testing Ability, you are instructed to compare his raw score to the entries in a table provided by the test publisher. This table gives the

_____ for the test.

9. After 8 months of school all the students at East Hills Junior High were given a test of paragraph comprehension. The mean raw score obtained by the 8th grade was 68. Considering only the students at East Hills, a raw score of 68 would correspond to a grade equivalent score of _____.

10. Thirty-five out of every forty college seniors score above the 55th percentile on a measure of computational ability. The norms of this measure were not computed from the scores of a representative sample of _____.

11. A test score indicates that Student X is at the 86th percentile in reading speed when compared to graduate students. This means that Student X scored (1) _____ than 14 percent of the

(2) _____ in the standardization sample from which

the test (3) _____ were computed.

12. A reporter investigating farm labor conditions in Midville, found that the DeMarco farm employed migrant workers at $1.00 per day and farm-machinery operators at $20 per day. The migrant workers made up 64% of the labor force while the farm-machinery operators made up 36% of those employed. The reporter concluded that the average wage paid to farm workers in Midville was $1.00 per day, and that legislation should be passed to increase farm wages.

    a. In computing the average salary, the reporter used what type of average score? _____

    b. What was the sample used to describe the average wage of farm workers in Midville? _____

In the following questions, circle the letter that corresponds to the phrase which best completes each question.

13. The average number of miles driven each day on a trip is computed by dividing the total distance traveled by the number of days on the road. The type of average used is a (circle one)

    a.) mean.
    b.) median.
    c.) neither of the above.

14. Before giving a test to a class, we know that the number of people who will score above the median for that class will be (circle one)

a) one-half the sum of the class scores.
b) one-half the number of class members.
c) about the same as those scoring above the mean.

15. Let us assume that urban populations score significantly different than rural populations on the Modern Vocabulary Test. In order for test results to be most useful, you should hope to find (circle one)

a) two sets of norms, one based on a rural population, another on an urban population.
b) one set of norms based on a population containing equal numbers of urban and rural subjects.
c) either of the above would be equally satisfactory.

17. A seventh grade teacher discovered that half her pupils scored below average on a standardized test of spelling in spite of her year-long efforts to teach them to spell. She should conclude that (circle one)

a) the wrong kind of average had been computed on the test.
b) her class was below average ability.
c) these results were what one might usually expect.

18. Many parents and teachers believe that all beginning 4th graders should have a grade equivalent score of 4.0 or better. Such a belief is absurd because (circle one)

a) not all children are alike - individual differences are important.
b) parents confuse the difference between mean and median scores.
c) by definition, approximately half the children will score below the mean.
d) class norms do not fully portray the range of abilities in a classroom of students.

19. Other things being equal, the usefulness of a test increases with (circle one)

a) an increase in the number of norms reported from different standardization samples.
b) an increase in the mean score.
c) a decrease in the number of cases reported for each standardization sample.

Questions 20 through 23 are based on the following paragraph taken from the Midville Daily News:

> "No wonder we are behind in the space race! The scores obtained by our own junior class at Midville High last Thursday are shocking! Fifty-five out of sixty juniors who took the "Modern Technology Test" scored lower than five percent. This shows conclusively that high school students all over the country are not being trained in science. The "Modern Technology Test" was developed and standardized on MIT graduates for selection in their new space-training program. If we are to get ahead in the space race, high school students must do better than this!"

20. In the above paragraph,

a) what was the sample? _____.

b) what was the population? _____.

c) what was the standardization sample reported for the "Modern Technology Test"?

_____.

21. The above report is not fair because it did not refer to (circle one)

a) more standardization samples of graduate students other than MIT students.
b) separate norms for men and women.
c) norms for high school students.
d) the mean score for MIT students so that grade equivalent scores could be assigned to the junior class.

22. If one wished to estimate how much science training high school students in the country are receiving, one would (circle one)

a) consult with prominent scientists.
b) do a more careful study of the junior class at Midville.
c) sample numbers of high school classes through the nation.
d) insist that every high school student in the country be tested.

23. Comparing the scores of high school juniors at Midville with the scores of graduate students at MIT is not helpful because (circle one)

a) most of the juniors do not wish to go to MIT when they graduate.
b) the science equipment is vastly different at both schools.
c) grade equivalent scores are not used beyond elementary school.
d) high school students are being compared with a group to which they do not belong.

24. Kathy, a beginning seventh grader, obtained a percentage score of 92 on the Reading Placement Test, which gave her a grade equivalent score of 8.0 and a percentile score of 67. In interpreting her test score, the teacher should inform Kathy's parents that (circle one)

a) Kathy is one year ahead of the 7th grade class in reading ability.
b) Kathy scored higher than two-thirds of her fellow 7th graders.
c) Kathy did very well by answering 92 percent of the items correctly on the test.

THE TIME AT THE COMPLETION OF THIS SECTION IS _____.